



HAL
open science

QEMSS: A selection scheme for participatory sensing tasks.

Rim Ben Messaoud, Yacine Ghamri-Doudane

► **To cite this version:**

Rim Ben Messaoud, Yacine Ghamri-Doudane. QEMSS: A selection scheme for participatory sensing tasks.. CFIP/NOTERE, Jul 2015, Paris, France. 10.1109/NOTERE.2015.7293493 . hal-01287687

HAL Id: hal-01287687

<https://hal.science/hal-01287687>

Submitted on 20 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

QEMSS: A selection scheme for participatory sensing tasks

Rim Ben Messaoud*[†] and Yacine Ghamri-Doudane[†]

*University of Paris-Est Marne La Vallée, Laboratory of Informatics Gaspard Monge (LIGM), Champs sur Marne, France

[†]University of La Rochelle, Laboratory of Informatics, Image and Interaction (L3I), La Rochelle, France

{rim.ben_messaoud, yacine.ghamri}@univ-lr.fr

Abstract—The new generation of smart devices, equipped with a large variety of sensors, enhances the Participatory Sensing of data. However, many issues arise when selecting participants to perform the sensing tasks. These issues are necessarily related to the limited energetic resources of devices, the impact of users mobility as well as the quality of collected data, recently defined as “Quality of Information” (QoI). In this context, we propose QEMSS (QoI and Energy aware Mobile Sensing Scheme) as a selection scheme for participatory sensing tasks, taking into consideration the quality of sensed data, QoI, and the dedicated energy for their acquisition. The aim of our model QEMSS is to select, among all participants in the sensing campaigns, the subset of users who maximizes the QoI of non redundant information while minimizing the overall energy consumption. To do so, we illustrate our selection scheme based on the Tabu Search algorithm in order to achieve a sub-optimal solution. Simulation results were compared to two other State of The art schemes: the Random Selection (RS) and a method based on a greedy search (DPS). Our scheme is proved to be as performing as the two other methods. Particularly, our scheme achieves a very high quality of information in challenging scenarios such as low dense areas and/or low energetic resources.

I. INTRODUCTION

Les nouveaux dispositifs mobiles sont non seulement dotés de capacités de connexion et de calcul, mais ils se présentent également avec une variété de capteurs embarqués tels que l’accéléromètre, le GPS, la caméra et le microphone [1]. Ceci appuie leur rôle dans la collecte et le partage des données exploitant la foule (Crowd), une forme de collecte participative appelée *Crowdsensing* [2]. Les recherches ont profité du potentiel de ce nouveau paradigme pour introduire plusieurs applications. A titre d’exemple, PEIR (Personnal Environment Impact report) est un système qui rassemble les données de localisation des mobiles et en génère des rapports de comportement urbain des participants [3]. D’autres applications prometteuses ont été développées dans différents domaines tels que la santé et la gestion du trafic routier [4].

Cependant, l’exploitation des dispositifs des individus lors de la collecte participative introduit de nouveaux défis. Un premier frein majeur pour l’acceptation de ces applications par les utilisateurs concerne la protection de leur vie privée. Dans ce but, Shin et al. ont présenté la plateforme *AnonymSense* [5] adressant l’anonymisation des informations privées des participants. Par ailleurs, et même en présence de mécanismes de confidentialité, les utilisateurs peuvent être réticents à consacrer leurs ressources, à moins de recevoir une indemnité adaptée. C’est dans ce contexte que les modèles de récompense dites “incentives” ont été introduits afin de contrôler le budget

des collectes de données à grande échelle. Citons l’algorithme *MCARD* [6] qui assure la sélection des participants au *Crowdsensing* tout en minimisant le coût total de leurs récompenses.

Dans un autre contexte, un intérêt particulier a été adressé au fait que les ressources énergétiques des terminaux mobiles restent limitées, ce qui nécessite des stratégies de sélection spécifiques pour réduire la quantité d’énergie prédestinée aux campagnes de collecte. Sheng et al. [7] ont montré que la collaboration entre les utilisateurs pouvait radicalement diminuer l’énergie globale consommée. De même, les auteurs de *StreamShaper* [8] ont proposé de faire des mesures périodiques d’une façon coopérative entre les utilisateurs pour réduire l’utilisation globale des batteries des dispositifs durant la collecte des données.

Tous les travaux précédents n’ont pas traité la notion de la qualité d’information, QoI, une exigence majeure du *Participatory Sensing*. Ce terme a été initialement présenté par Bisdikan et al. pour les réseaux de capteurs [9] puis récemment étudié pour le *Crowdsensing*. Cependant, le travail présenté par les auteurs de [10] a ciblé la qualité des données extraites en attribuant des récompenses aux utilisateurs en fonction de leur participation. Cet article illustre un algorithme de sélection des participants pour maximiser la satisfaction des requêtes en termes de QoI sous contrainte budgétaire. Par ailleurs, la problématique de l’énergie a été partiellement abordée.

En considérant les challenges de la collecte participative cités ci-dessus, nous proposons QEMSS (QoI and Energy-aware Mobile Sensing Scheme) notre modèle de déploiement de la collecte participative attentif à la QoI et l’énergie introduit aussi dans [11]. L’objectif de notre approche consiste à maximiser la qualité globale des données extraites tout en respectant les contraintes énergétiques des dispositifs des participants. Nous modélisons d’abord le problème d’optimisation qui correspond à notre requête et nous proposons un algorithme qui se base sur la recherche tabou pour atteindre une solution sous-optimale. L’efficacité de notre stratégie a été intensivement évaluée en termes du taux maximal de QoI réalisée, ainsi que la précision spatiale et temporelle des données. Bien que les résultats pour les régions de collecte à forte densité sont comparables à d’autres méthodes de sélection de l’état de l’art, l’importance de notre modèle a été démontrée pour des régions à faible densité en participants et pour sa flexibilité en termes de ressources nécessaires pour la collecte.

Nous abordons cet article par présenter succinctement en section II la technique utilisée dans ce travail: la recherche tabou. Notre modèle est ensuite détaillé avec les éléments

nécessaires pour la modélisation du problème d'optimisation qui lui correspond dans la section III. La section IV explique l'exploitation de la recherche tabou pour concevoir notre solution QEMSS. Les résultats de la simulation sont illustrés dans la section V où QEMSS est comparé à deux méthodes concurrentes: DPS [10] et la sélection aléatoire. Finalement, les conclusions de ce travail et les perspectives envisagées sont présentées en section VI.

II. TECHNIQUE UTILISÉE: LA RECHERCHE TABOU

La recherche tabou (Tabu Search) est une méta-heuristique qui explore l'espace des solutions au-delà de l'optimalité locale. Cette technique a été introduite par Fred Glover en 1986 [12]. L'idée principale de la recherche tabou consiste à interdire des mouvements qui font retourner à des solutions visitées d'où le terme solutions taboues.

Soit S l'ensemble des solutions possibles d'un problème. Pour chaque solution $s \in S$, il existe un sous-ensemble de S appelé voisinage (Neighborhood) de s , $N(s)$. Le voisinage contient des solutions possibles obtenues en appliquant un mouvement simple (transformation) à s . Un mouvement m consiste à changer ou remplacer un élément de la solution actuelle. De plus, l'algorithme de la recherche tabou (TS) utilise une structure de mémoire appelée liste tabou (Tabu List) pour éviter les cycles. Une solution appartenant au voisinage $N(s)$ n'est sélectionnée que si elle ne figure pas dans la TL. À chaque itération, le TS met à jour la liste tabou (TL) en y ajoutant les attributs des solutions visitées. En générale, les attributs ne contiennent pas la solution complète pour faciliter l'utilisation de la TL. La taille de la liste tabou, appelée *Tabu Tenure* (TT), est le nombre d'itérations pendant lesquelles un mouvement est déclaré comme tabou.

La recherche tabou débute par une phase d'initialisation pour générer une première solution possible s_{init} . Notons que le nombre d'itérations de l'algorithme TS dépend fortement de la nature de cette solution initiale. Le plus loin est s_{init} du voisinage de la solution sous-optimale, le plus important serait le temps d'exécution de la méthode. Par conséquent, il est plus convenable de ne pas commencer par une solution aléatoire.

L'utilisation de cette technique pour la conception de notre modèle de déploiement de collecte participative, QEMSS, est détaillée en section IV.

III. LE SYSTÈME PROPOSÉ ET LA FORMULATION DU PROBLÈME D'OPTIMISATION

A. Le système proposé

Nous considérons un ensemble N de participants et un serveur centralisé dans le cloud comme l'indique la figure 1. Chaque utilisateur, muni de son dispositif mobile équipé de capteurs embarqués, est considéré comme un participant candidat au *Crowdsensing*. Les utilisateurs se déplacent arbitrairement dans une région 2D modélisant une ville. Néanmoins, ils ne se déplacent pas d'une manière totalement aléatoire mais ils suivent plutôt des chemins. Par conséquent, la trajectoire de chaque utilisateur peut être représentée par l'ensemble des positions visitées. Par conséquent, nous subdivisons la région de collecte en sous-régions d'une ou plusieurs positions.

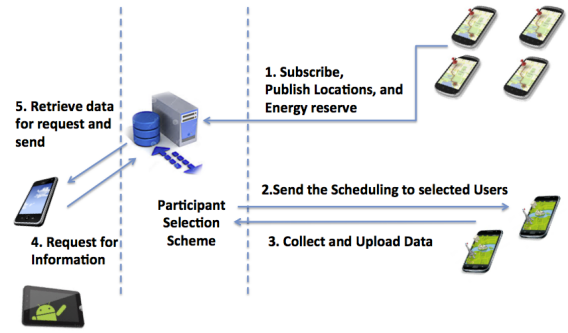


Fig. 1. Scénario du Mobile Crowdsensing: Le serveur sélectionne un sous-ensemble d'utilisateurs qui maximise la qualité des données puis diffuse le plan de la collecte. Les participants rassemblent et partagent les mesures. Finalement, le serveur répond aux requêtes d'applications tierces par les informations correspondantes.

En début de chaque période de sensing, le serveur initialise les tâches en précisant les exigences en qualité de données, telles que le nombre d'échantillons, la région et la période de collecte. Le serveur reçoit périodiquement les coordonnées des utilisateurs souscrits au *Crowdsensing* en plus de l'état de leurs ressources énergétiques. Notons que nous considérons dans ce travail que la consommation unitaire en énergie d'un capteur, notée e_c , est la même pour différents modèles de mobiles [13]. De plus, nous introduisons un seuil énergétique e_θ que lorsque le niveau de batterie d'un terminal atteint, l'utilisateur correspondant est éliminé de la liste des candidats aux futures tâches de collecte. Ceci garantit aux participants l'énergie nécessaire à l'utilisation quotidienne de leurs mobiles.

Le serveur sélectionne ensuite le sous-ensemble d'individus qui maximise la qualité globale des données collectées en considérant le coût énergétique de leur acquisition ainsi que la redondance de l'information. Pour ce faire, nous limitons le nombre des mesures exigées à une seule mesure par région pour une période de collecte précise. Finalement, le plan de déploiement des tâches de collecte est diffusé pour annoncer à chaque participant où et quand intervenir.

B. Formulation du problème d'optimisation

1) *Les attributs et les métriques de QoI*: La QoI peut être décrite par des attributs comme la précision, l'exactitude, la validité et la fiabilité [14]. Dans notre travail, nous considérons trois attributs pour qualifier les données extraites: la *completeness*, la *timeliness* et l'*affordability*. Le premier terme décrit si la donnée extraite donne une vision complète sur la réalité. Nous étudions la *completeness* spatiale en estimant la zone potentiellement couverte par un participant. La *timeliness* exige que la collecte et le partage des données se déroulent dans un intervalle du temps précis. Le dernier critère souligne le fait qu'un individu muni d'un dispositif de faible batterie ne peut pas participer aux campagnes de collecte.

Afin d'estimer la qualité des données qu'un sous ensemble X d'utilisateurs peut fournir, nous quantifions les différents attributs mentionnés ci-dessus par la fonction d'utilité proposée par les auteurs de [15] comme métrique. La fonction d'utilité a été introduite en économie pour décrire la valeur d'un produit perçue par un client puis utilisée dans différents domaines

comme fonction de normalisation. Ainsi, nous utilisons des “utilités” différentes pour quantifier chaque attribut de QoI de point de vue système.

Comme l’indique la revue de la littérature [15], la *completeness* est un critère ascendant puisqu’un utilisateur rassemble plus d’échantillons lorsqu’il couvre plus d’espace. Prenant le cas où la couverture x varie comme suit, $x_\alpha \leq x \leq x_\beta$ avec x_m la médiane de cette variation, la fonction d’utilité correspondante est:

$$U_c(x) = \begin{cases} 0 & x < x_\alpha & (1a) \\ \frac{\left(\frac{x-x_\alpha}{x_m-x_\alpha}\right)^\zeta}{1 + \left(\frac{x-x_\alpha}{x_m-x_\alpha}\right)^\zeta} & x_\alpha \leq x \leq x_m & (1b) \\ 1 - \frac{\left(\frac{x_\beta-x}{x_\beta-x_m}\right)^\gamma}{1 + \left(\frac{x_\beta-x}{x_\beta-x_m}\right)^\gamma} & x_m < x \leq x_\beta & (1c) \\ 1 & x > x_\beta, & (1d) \end{cases}$$

où

$$\zeta \geq \max\left\{\frac{2(x_m - x_\alpha)}{x_\beta - x_m}, 2\right\}. \quad (2)$$

et

$$\gamma = \frac{\zeta(x_\beta - x_m)}{x_m - x_\alpha}. \quad (3)$$

ζ et γ sont les paramètres de la pente.

Pour la *timeliness*, nous proposons une fonction centrée autour d’un intervalle de temps avec l’instant de mesure comme médiane. Ceci assure que des mesures collectées à des instants différents du moment de collecte désiré auront une utilité faible. De même, nous considérons la variation du temps x , $x_\alpha \leq x \leq x_\beta$ et le moment de collecte désiré x_d , et nous formulons la fonction d’utilité décrite dans l’équation (4).

$$U_t(x) = \begin{cases} 0 & x < x_\alpha, x > x_\beta & (4a) \\ \frac{\left(\frac{x-x_\alpha}{x_{m1}-x_\alpha}\right)^\zeta}{1 + \left(\frac{x-x_\alpha}{x_{m1}-x_\alpha}\right)^\zeta} & x_\alpha \leq x \leq x_{m1} & (4b) \\ 1 - \frac{\left(\frac{x_d-x}{x_d-x_{m1}}\right)^\gamma}{1 + \left(\frac{x_d-x}{x_d-x_{m1}}\right)^\gamma} & x_{m1} < x \leq x_d & (4c) \\ 1 - \frac{\left(\frac{x-x_d}{x_{m2}-x_d}\right)^\zeta}{1 + \left(\frac{x-x_d}{x_{m2}-x_d}\right)^\zeta} & x_d < x \leq x_{m2} & (4d) \\ \frac{\left(\frac{x_\beta-x}{x_\beta-x_{m2}}\right)^\gamma}{1 + \left(\frac{x_\beta-x}{x_\beta-x_{m2}}\right)^\gamma} & x_{m2} < x \leq x_\beta & (4e) \end{cases}$$

où x_{m1} et x_{m2} sont les médianes respectives des intervalles de temps $[x_\alpha, x_d]$ et $[x_d, x_\beta]$.

Finalement, l’utilité totale de la donnée extraite est le produit des utilités de ses attributs de QoI présenté par l’équation (5). Cette formulation souligne le fait qu’une information est “utile” seulement si elle répond à tous les critères de qualité désirée. Par conséquent, si une quantité est nulle, la totalité de l’utilité de donnée serait nulle. Ceci permet d’éliminer les mesures hors la zone ou la période du *sensing*.

$$U_{QoI} = U_c(x) \times U_t(x). \quad (5)$$

D’une part l’équation (5) illustre la fonction objectif de notre modèle où nous voulons maximiser la *timeliness* et la

completeness des données extraites du *Crowdsensing*. D’autre part, la métrique qui quantifie l’attribut de l’*affordability* serait introduite comme une contrainte dans notre problème d’optimisation vu que c’est une condition limitant l’accès aux tâches. Ce dernier attribut est un critère descendant [15]. L’utilité des mesures d’un dispositif diminue en fonction de ses ressources. Par conséquent, nous mesurons l’accessibilité d’un terminal à la collecte participative par la fonction de l’équation (6).

$$U_e(x) = 1 - U_c(x). \quad (6)$$

Notons que le taux de l’énergie x varie dans $[e_\theta, 100\%]$ où $U_e(e_\theta) = 0$.

2) *Le problème d’optimisation*: La finalité de notre modèle QEMSS consiste à maximiser la qualité des données collectées par le sous ensemble X des participants sélectionnés tout en diminuant le coût de leur acquisition en termes de ressources énergétiques. Nous proposons la formulation suivante:

$$\begin{aligned} & \underset{x}{\text{maximize}} && \sum_{i=1}^N \sum_{j=1}^M x_{ij} \times U_{QoI}(QoI_{ij}) \\ & \text{subject to} && U_e(e_i - e_c \sum_{j=1}^M x_{ij}) \geq 0 \quad \forall i \in 1 \dots N \quad (7) \\ & \text{and} && \sum_{i=1}^N x_{ij} \leq 1 \quad \forall j \in 1 \dots M \end{aligned}$$

X est une matrice de dimension *users* \times *areas* dont les éléments sont dans l’ensemble $\{0, 1\}$. Si la valeur de x_{ij} est égale à 1, l’utilisateur i est sélectionné pour rassembler des mesures dans la région j . La première contrainte du problème (7) assure pour chaque terminal l’énergie nécessaire à ses activités quotidiennes pendant que la deuxième contrainte interdit la redondance de l’information en limitant le nombre d’utilisateurs sélectionnés par sous-région à 1.

Dans ce travail, nous cherchons la “combinaison” des utilisateurs qui répond aux contraintes et maximise la fonction objectif du problème (7). Néanmoins, les problèmes d’optimisation combinatoire sont difficiles voire impossibles à résoudre lorsqu’il s’agit d’un nombre important d’éléments impliqués. En effet, nous proposons ici un modèle de déploiement de campagnes de collecte à grande échelle, ce qui complique la résolution de ce problème [10]. En conséquence, il serait plus efficace de faire appel à un algorithme méta-heuristique tel que la recherche tabou (Tabu Search) pour générer une solution sous-optimale.

IV. QEMSS: QOI AND ENERGY-AWARE MOBILE SENSING SCHEME

Dans cette section, nous détaillons l’exploitation de la recherche tabou (TS) pour concevoir notre solution QEMSS. Nous présentons les composantes du TS décrites en section II liées aux éléments de notre système :

- Une solution envisageable X , décrite ci-dessous, est une matrice booléenne de taille *users* \times *areas*: x_{ij} est dans $\{0, 1\}$. Cette solution doit répondre aux contraintes du problème (7). En premier lieu, l’énergie résiduelle d’un dispositif mobile, $e_i - e_c \sum_{j=1}^M x_{ij}$,

doit être supérieure ou égale au seuil énergétique prédéfini e_θ . De plus, nous affectons un seul participant à chaque sous-région. En d'autres termes, la somme de chaque colonne a_j de la solution X doit être inférieure ou égale à 1.

$$X = \begin{matrix} & a_1 & \dots & a_M \\ u_1 & \left(\begin{matrix} x_{11} & \dots & x_{1M} \\ \vdots & x_{ij} & \vdots \\ x_{N1} & \dots & x_{NM} \end{matrix} \right) \\ \vdots & & & \\ u_N & & & \end{matrix},$$

- Un mouvement $m \in M(X)$ est une modification appliquée à la solution actuelle X pour en générer d'autres solutions possibles. Nous considérons comme mouvement la permutation d'attribution des tâches entre deux participants qui coexistent dans une même zone à couvrir. Par conséquent, le mouvement m peut être présenté par une matrice de même taille que la solution X avec tous ses éléments nuls sauf les deux positions (utilisateurs) à permuter.
- Pour générer le voisinage d'une solution X , nous appliquons le mouvement m que nous avons défini. Chaque élément du voisinage $N(X)$ est alors le résultat d'une simple opération ou exclusif (XOR): $X' = X \oplus m$.
- Nous calculons, pour chaque solution $X \in S$, le taux maximal de la qualité des données réalisée et nous le considérons comme son attribut principal.
- À chaque itération, nous mettons à jour la liste tabou (TL) en ajoutant les attributs des meilleures solutions visitées X_{best} en termes de QoI.

En choisissant le taux maximal de QoI comme étant l'attribut de X , nous interdisons non seulement le retour vers une même solution mais aussi de ne pas considérer les solutions visitées avec une même valeur de QoI, technique adoptée par les auteurs du travail [16]. Par ceci, nous réduisons le temps de calcul et l'espace de mémoire nécessaire à la sauvegarde de la liste tabou (TL).

Dans ce qui suit, nous illustrons les différentes étapes de notre solution QEMSS basée sur la recherche tabou pour la résolution du problème d'optimisation (7) et présentée par l'**algorithme 1**.

Initialisation: Cette première phase vise à générer une solution initiale, X_{init} , qui répond aux contraintes de notre problème. Cependant, il convient à mentionner qu'une solution risque d'attribuer des tâches de collecte à des utilisateurs qui n'existent pas en réalité dans la région en question. De ce fait, nous utilisons une matrice référence X_{ref} , avec $x_{ref}(i, j) = 1$ si et seulement si la position j appartient à la trajectoire de l'utilisateur i . Puis, nous générons une première solution en se basant sur cette matrice. L'heuristique présentant cette première phase de TS est décrite par l'**algorithme 2**.

Formation du voisinage: À partir d'une solution X , nous pouvons générer tout son voisinage $N(X)$ en appliquant un seul mouvement m . Comme nous l'avons proposé ci-dessus, nous considérons comme mouvement l'échange d'affectation d'une tâche de collecte dans une zone déterminée entre deux

Algorithm 1 TS pour maximiser la QoI des données collectées

- 1: **Initialisation:** Générer une solution initiale X_{init} dans l'espace des solutions envisageables S et calculer son utilité en termes de QoI, U_{init} .
 - 2: $X \leftarrow X_{init}$
 - 3: $X_{opt} \leftarrow X_{init}$
 - 4: $U_{opt} \leftarrow U_{init}$
 - 5: **while** $iter < Max_{iter}$ **do**
 - 6: **Formation du voisinage:** Générer tout le voisinage de la solution courante X en appliquant tous les mouvements $m \in M(X)$ sauf ceux considérés comme tabous.
 - 7: **Sélection du meilleur voisin:** Choisir la solution voisine $X' \in N(X)$ avec la meilleure valeur de QoI calculée, U_{best} , comme solution initiale de l'itération suivante.
 - 8: **if** $U_{best} > U_{opt}$ **then**
 - 9: $X_{opt} \leftarrow X_{best}$
 - 10: $U_{opt} \leftarrow U_{best}$
 - 11: **end if**
 - 12: **Mise à jour de TL:** Ajouter les attributs des solutions visitées à la liste tabou (TL).
 - 13: $X \leftarrow X_{best}$
 - 14: **end while**
 - 15: Retourner X_{opt}
-

utilisateurs présents. Pour cela, nous nous basons sur la matrice X_{ref} . Ainsi, les futures attributions sont choisies parmi les éléments unitaires d'une colonne (région) de la matrice X_{ref} . Cette méthode réduit clairement la taille du voisinage à explorer pour trouver la solution optimale et par conséquent le temps d'exécution de notre algorithme de recherche basé sur la recherche tabou (TS).

Algorithm 2 Méthode Greedy pour générer une solution initiale

Require: Ensemble de sous-régions de collecte R , Utilisateurs N , Matrice d'utilité U .

Ensure: Le sous-ensemble d'utilisateurs X

- 1: **while** R n'est pas vide **do**
 - 2: Sélectionner aléatoirement une zone $R_i \in R$
 - 3: **while** $select = 0$ **do**
 - 4: Sélectionner l'utilisateur qui fournira la valeur maximale de QoI dans la sous-région R_i :
 $(Max_U, argmax) \leftarrow \max(U_{R_i})$
 - 5: Vérifier la contrainte énergétique:
 - 6: **if** $U_e(argmax) > 0$ **then**
 - 7: $X(argmax, R_i) \leftarrow 1$
 - 8: $select \leftarrow 1$
 - 9: **else**
 - 10: $U(argmax, R_i) \leftarrow 0$
 - 11: $select \leftarrow 0$
 - 12: **end if**
 - 13: **end while**
 - 14: $R \leftarrow R - R_i$
 - 15: **end while**
 - 16: Retourner X
-

V. ÉVALUATION DES PERFORMANCES

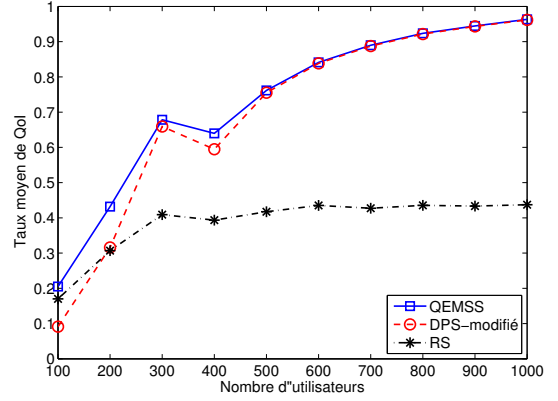
Nous articulons notre simulation autour d'une modélisation d'une zone de collecte de taille $4000m \times 6000m$. Cette zone représente une région urbaine subdivisée en une vingtaine de rues horizontales et autres verticales (20). Cette topologie est similaire au modèle répandu de la cité Manhattan (New York). Par conséquent, nous utilisons des traces de mobilité du modèle de mobilité Manhattan (Manhattan mobility model) généré par le simulateur des traces, Bonnmotion [17]. Chaque participant entre un chemin avec une vitesse aléatoirement choisie dans l'intervalle $[1, 3] m.s^{-1}$, puis continue son mouvement tout droit avec une probabilité égale à 0.5, ou encore il tourne à gauche ou à droite avec des probabilités égales à 0.25 chacune. Nous faisons varier le nombre des participants présents dans la région de collecte entre 100 et 1000. En outre, l'énergie initiale des dispositifs de chaque utilisateur est arbitrairement générée suivant une loi uniformément distribuée entre e_θ et 100%. Au total, nous avons réalisé des simulations pour une période de collecte de 16 heures par jour pour différents nombres d'individus impliqués aux campagnes du *Crowdsensing*.

Nous présentons les résultats obtenus liés principalement à la valeur maximale de la qualité des données (QoI) réalisée en fonction du temps et du nombre de participants. De plus, nous détaillons les performances de notre système en termes des précisions spatiale et temporelle des données extraites. QEMSS a été comparé à une version modifiée de l'algorithme *Dynamic Participant Selection Scheme* (DPS) introduit dans [10] et à une méthode de sélection aléatoire comme référence. Les résultats sont illustrés dans les figures 2 et 3.

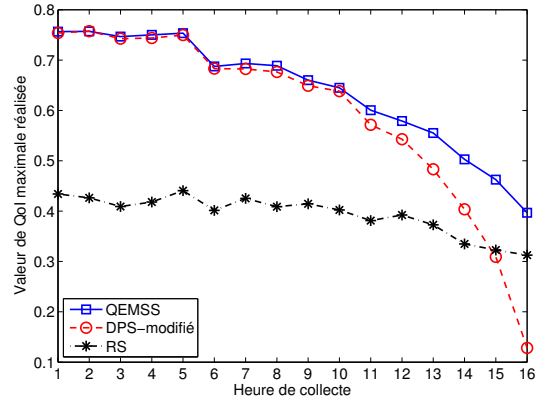
A. Valeur maximale de QoI

L'illustration 2 représente la valeur maximale de la qualité des données collectées, mesurée par l'équation (5) décrite en section III. L'évolution du taux moyen de la QoI, atteint durant la totalité de la période de collecte, en fonction du nombre des participants est tracée dans la figure 2-a. Bien évidemment, le taux de la QoI réalisée augmente avec le nombre d'utilisateurs présents dans la zone à couvrir. Particulièrement, les trois approches comparées fonctionnent aussi efficacement dans les régions à densité importante (500-1000 participants). Néanmoins, notre solution QEMSS est nettement plus performante que les deux modèles, DPS et RS dans les régions à faible densité d'utilisateurs (100-400). Ceci est dû essentiellement au fait que la recherche tabou que nous effectuons tout au long de la sélection varie les utilisateurs désignés pour les requêtes. Par conséquent, nous n'abusons pas des ressources énergétiques des participants et nous pouvons choisir, ainsi ceux qui fournissent la meilleure qualité de donnée.

La figure 2-b décrit la variation de la valeur maximale de la QoI réalisée par 400 participants durant une période de collecte de 16 heures. Nous observons qu'aux premières heures de mesure, la performance de notre modèle est comparable à celle de l'algorithme DPS. Par contre, l'écart devient de plus en plus important vers la fin de la période jusqu'à atteindre 27%. Ceci résulte du type de la recherche appliquée dans les deux modèles. La recherche tabou (TS), qu'utilise notre solution QEMSS, explore intensivement le voisinage des solutions pour diversifier les utilisateurs sélectionnés à chaque itération.



(a) Le taux moyen de la QoI en variant le nombre des participants



(b) La valeur maximale de la QoI réalisée par 400 participants durant une période de collecte de 16-heures

Fig. 2. La valeur de la QoI réalisée par différents modèles de sélection de participants en variant (a) le nombre des utilisateurs dans la zone à couvrir (b) et le temps de collecte des mesures pour 400 participants.

Cependant, la recherche glouton (Greedy search) exploitée par l'algorithme DPS attribue aux mêmes participants plusieurs tâches, ce qui draine l'énergie de leurs dispositifs et les exclue de la liste des candidats. D'autre part, la sélection aléatoire a clairement la performance la plus faible en termes de QoI.

B. Précision spatiale et temporelle

Pour mesurer la précision spatiale des données (spatial accuracy), nous utilisons la distance de Manhattan décrite dans l'équation (8). Nous calculons alors la distance entre les coordonnées de la zone de collecte désirée, annoncée au début des campagnes par le serveur, et les coordonnées spatiales de la donnée extraite en fin de la période de collecte. Notons que pour répondre aux requêtes lancées de la part des applications tierces, le serveur cherche les mesures les plus adéquates en termes de précision spatiale et temporelle. Par conséquent, la plus importante est la distance de Manhattan $dist(area_{req}, data_{coll})$, la moins précise est l'information collectée à ce niveau.

$$s_a = 1 - \sum_{i=1}^n |X_i - Y_i|, \quad (8)$$

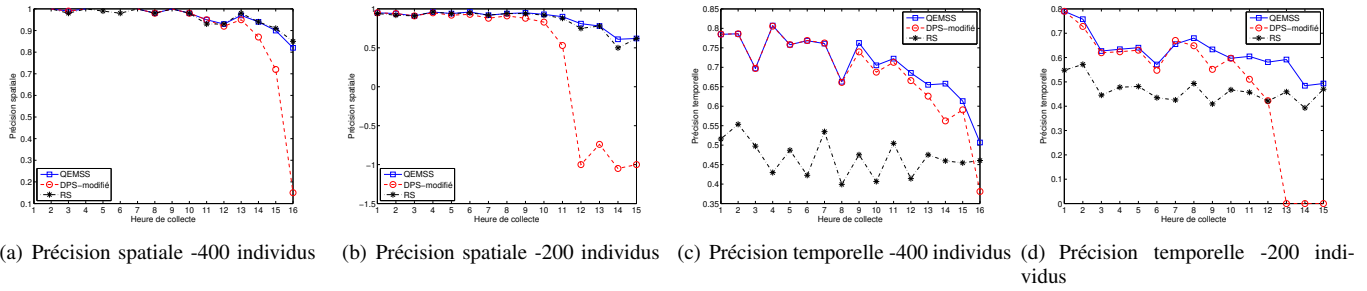


Fig. 3. Précision spatiale (a) et (b), et temporelle (c) et(d) des données pour différentes densités d'utilisateurs (200-400)

Comme le montre les figures 3(a) et 3(b), la précision des données collectées au sein du modèle DPS diminue en fonction du temps. Ceci se conforme avec les résultats illustrés ci-dessus concernant la non efficacité de cet algorithme à la fin de la période du sensing en termes de QoI. Toutefois, notre solution QEMSS ainsi que la sélection aléatoire (RS) accomplissent une précision spatiale plus importante et similaire tout au long de la période du sensing.

Les figures 3(c) et 3(d) illustrent l'évolution de la précision temporelle des algorithmes en fonction du temps. Cette mesure quantifie l'attribut de la *timeliness* des données par la fonction d'utilité $U_t(x)$ prédéfinie en section III. Nous remarquons alors que le modèle aléatoire est le moins performant à ce niveau. En effet, RS sélectionne des participants pour couvrir plus des zones sans prendre en considération le critère des instants de mesure. D'autre part, et bien que l'algorithme DPS était aussi précis que notre solution pour la première demi-période de la collecte, son taux d'erreur s'élèvera à 62% vers la fin. Ceci est dû à la limitation des ressources énergétiques des participants à la fin d'une période du sensing.

VI. CONCLUSIONS

Dans cet article, notre objectif est de présenter un modèle de déploiement des tâches de collecte qui permet de faire un compromis entre l'énergie consommée par les applications mobiles du sensing et la qualité des données extraites. Nous proposons ici la quantification des différents attributs de la QoI en attribuant à chaque critère une fonction d'utilité adéquate. Pour atteindre notre finalité, nous avons basé notre solution sur l'algorithme de la recherche tabou (TS). Les résultats des simulations démontrent que notre système peut atteindre une performance compétitive dans les régions de forte population quand comparé à deux autres méthodes de l'état de l'art. Cependant, nous réalisons un gain significatif en termes de QoI et de précision spatiale et temporelle notamment dans les scénarios difficiles tels que les zones à faible densité ou encore les participants munis de dispositifs à faibles ressources énergétiques.

Pour conclure, notre solution a une meilleure performance que les deux modèles de déploiement de collecte participative; RS et DPS. Néanmoins, il demeure important que nous assurions non seulement la minimisation du coût global du *Crwodensing* mais de créer également un système équitable vis à vis aux participants.

REFERENCES

- [1] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.
- [2] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *WSW: Mobile Device Centric Sensor Networks and Applications*, 2006, pp. 117–134.
- [3] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research," in *Mobile Systems, Applications and Services*, 2009, pp. 55–68.
- [4] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *ACM conf. on Embedded Network Sensor Systems*, 2008, pp. 323–336.
- [5] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, "AnonySense: Privacy-aware people-centric sensing," in *Mobile Systems, Applications and Services*, 2008, pp. 211–224.
- [6] X. Lu, D. Li, B. Xu, W. Chen, and Z. Ding, "Minimum cost collaborative sensing network with mobile phones," in *International Conference on Communications (ICC)*, 2013, pp. 1816–1820.
- [7] X. Sheng, J. Tang, and W. Zhang, "Energy-efficient collaborative sensing with mobile phones," in *INFOCOM*, 2012, pp. 1916–1924.
- [8] H. Weinschrott, F. Durr, and K. Rothermel, "Streamshaper: Coordination algorithms for participatory mobile urban sensing," in *Mobile Adhoc and Sensor Systems (MASS)*, 2010, pp. 195–204.
- [9] C. Bisdikian, J. Branch, K. Leung, and R. Young, "A letter soup for the quality of information in sensor networks," in *Pervasive Computing and Communications (PerCom)*, 2009, pp. 1–6.
- [10] Z. Song, C. Liu, J. Wu, J. Ma, and W. Wang, "QoI-aware multitask-oriented dynamic participant selection with budget constraints," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4618–4632, 2014.
- [11] R. Ben Messaoud and Y. Ghamri-Doudane, "QoI and energy-aware mobile sensing scheme: A tabu-search approach," in *IEEE 82nd Vehicular Technology Conference: VTC2015-Fall*, 2015.
- [12] F. Glover, "Tabu search-Part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.
- [13] I. Koenig, A. Q. Memon, and K. David, "Energy consumption of the sensors of smartphones," in *Int. Symposium on Wireless Communication Systems (ISWCS)*, 2013, pp. 1–5.
- [14] V. Sachidananda, A. Khelil, and N. Suri, "Quality of information in wireless sensor networks: A survey," in *International Conference on Information Quality (ICIQ)*, 2010.
- [15] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," in *IEEE Network Operations and Management Symposium, NOMS*, 2008, pp. 144–151.
- [16] H. Kamal, M. Coupechoux, and P. Godlewski, "A tabu search DSA algorithm for reward maximization in cellular networks," in *Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2010, pp. 40–45.
- [17] N. Aschenbruck, R. Ernst, E. Gerhards-Padilla, and M. Schwamborn, "Bonnmotion: a mobility scenario generation and analysis tool," in *Int. Conf. on Simulation Tools and Techniques (ICST)*, 2010.